

29.8.21

נתוני עתק ומחשוב ענן

פרויקט סיום

כללי

מערכות לעיבוד נתוני עתק משמשות כלי רב עוצמה לניהול סיכונים, הפקת תובנות עסקיות וגיבוש אסטרטגיה בארגונים. הצורך בפיתוח ועיצוב מערכות מסוג זה תופסת תאוצה ע"י שילוב משאבי On-premise עם שירותי ענן מספקים מובילים. מערכות נתוני עתק מאפיינות חברות דיגיטליות המנצלות את עידן האינטרנט לביצוע משימות ליבה באמצעים שאינם דורשים דיגיטציה מיוחדת לתהליכים, אלא מהווים חלק מהם: לגוגל, פייסבוק וחברות כמותן אין קיום שלא בעולם הסייבר ומלכתחילה כל פעילות מתועדת וממוחשבת - אי לכך המעבר לניצול נתונים תפעוליים לצורך קבלת החלטות וביצוע Big Data Analytics הוא טבעי, מאחר והתמודדות עם אתגרי נפחים ועומסים היא חלק אינטגרלי מסדר יומם הטכנולוגי. בשנים האחרונות גם חברות העוסקות בפעילות ריאלית לא וירטואלית רותמות טכנולוגיות וכלים מתחום נתוני עתק. בפרויקט זה נעצב מערכת העוקבת אחר משלוחי חבילות, מרגע עלייתן למטוס ועד לקליטתן במרכז לוגיסטי של חברת השילוח בארץ, בארכיטקטורה אופיינית המשתמשת בתשתיות נתוני עתק מקובלות לניתוחם באופן מהיר ואדפטיבי.

מטרה

עיצוב מערכת אנליטיקות בתפיסת נתוני עתק הרוותמת NoSQL Databases ותפיסה מונחית שירותים (לוקאליים ומבוססי ענן) בארכיטקטורה טיפוסית ומימוש אתגרי אינטגרציה תשתיות.

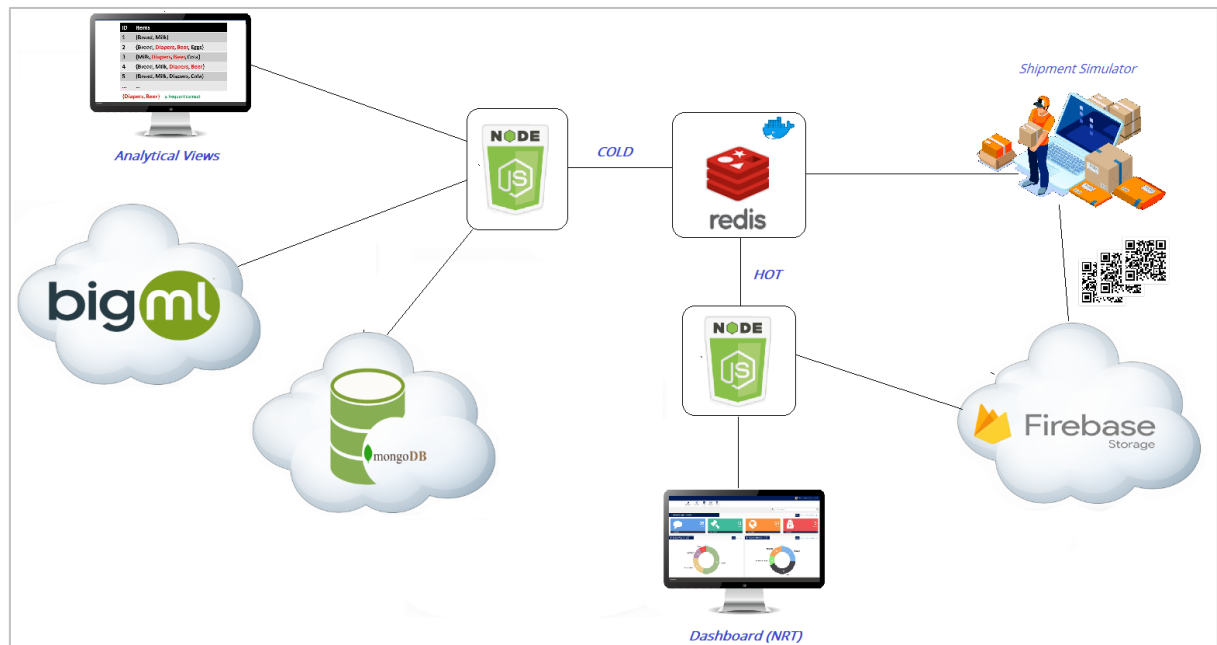
דרישות פונקציונאליות

- מאפייני חבילה כוללים:
 - מספר מעקב
 - רשימת פריטים
 - גודל (קטן, גדול, בינוני)
 - חיוב מס (עד \$75 פטור ממסים, עד \$500 חייב מע"מ, מעל חייבים במע"מ, מס קניה ומכס)
 - כתובת
 - מחוז יעד
- המערכת תאפשר צפייה במאפייני חבילות הנמצאות בדרכן לארץ: התצוגה תאורגן בחתך מחוז, תציג מספר חבילות בדרכן וכן ניתן לבקש גראף המציג התפלגות גדלים ואופי חיוב במס למחוז מסוים.
- סימולטור ידמה סריקת חבילות לפני משלוחן לארץ ויעדן יציאתה באמצעות מסר. כדי לדמות הגעה יפיק הסימולטור באופן אקראי מדבקת QR CODE אותה ישמור בשירות אחסון בענן (נא לא להשתמש ב- Real Time Database).
- מסר המגיע מהסימולטור מעדכן את מספר ואופי החבילות שבדרך.
- סקירת איזור האחסון באינטרוול זמן קונפיגורטיבי, שתמצא ותפענח את תמונות ה- QR CODE, תדמה הגעת המשלוח לארץ ותעדכן את מספר ואופי החבילות שעדיין לא הגיעו.
- המערכת תאפשר למידת חוקיות שכיחות (Frequent Item Sets) מתוכן החבילות ותצוגתן בצורת טבלה או דיאגרמה מתאימה לרבות חישוב מדדי Support Count & Support.

דרישות לא פונקציונאליות

- המערכת תעוצב בגישה מונחית Micro-Services.
- למערכת שלוש תת מערכות. כל תת מערכת תעוצב באמצעות תבנית MVC.
- עיצוב המערכת הכללי יתבסס רעיונית על תבנית Lambda architecture.

- הבסיס לשרתי האפליקציה ולתצוגה יהיו Node.js/Express וחבילות הרחבה מתאימות.
- המערכת תשתמש בשני מסדי נתונים מסוג MongoDB: NoSQL ישמור נתונים לשימושים היסטוריים ארוכי טווח, והאילו Redis ישמור נתונים מעודכים אד-הוק וזמינים לשאילתות מהירות.
- שרת Redis יישמש גם כ- Message Broker, אופציונאלית ניתן לבחור ב- Apache Kafka.
- יש להשתמש בשירות DBaaS עבור MongoDB (כגון ATLAS : <https://www.mongodb.com/cloud/atlas>).
- יש להשתמש בשרת Redis באמצעות Docker Image (כגון [Redis - Official Image | Docker Hub](https://hub.docker.com/_/redis)).



דיאגרמה 1: מיפוי טכנולוגי מוצע למערכת מעקב וניתוח משלוחי חבילות

דגשים:

- מועד הגשה נוכחי : 21.10.21 (יתכנו שינויים, נא לעקוב אחר מודל).
- יש להרשם בלינק הבא לשעת הגשה: [הגשת פרויקט סיום אריאל קיץ 2021 נתוני עתק Google Sheets](#)
- את הפרוייקט מגישים קבוצות של שלושה סטודנטים.
- סטודנט יציין בהגשה מה היתה אחריותו ומה מבין המודולים השונים הוא המבצע ועליהם יישאל.
- במהלך ההגשה יישאלו שאלות על החומר הנלמד.

משאבי עזר:

[Add Firebase to your JavaScript project \(google.com\)](#)

[Associations | BigML.com API](#)

[Association Rule Mining: What Frequent Itemsets is all about | by Salil Jain | Towards Data Science](#)

[Complete guide to Association Rules \(1/2\) | by Anisha Garg | Towards Data Science](#)

[Microsoft PowerPoint - AssociationAnalysis-Part1.pptx \(wmich.edu\)](#)

בהצלחה.

